09 | 性能测试的种类：如何快准狠地抓住一个测试的本质？

2019-12-16 庄振运

讲述：庄振运
时长 14:40 大小 13.45M

你好，我是庄振运。

从这一讲开始我们讨论性能测试。性能测试是一种特殊的软件测试，它的目的是**确保软件应用程序在一定的负载流量下运行良好**。性能测试是性能分析和性能优化的基础，它的目标是**发现和性能相关的各种问题和性能瓶颈，从而进一步去消除错误和性能瓶颈**。

由于性能测试本身就有好多种类；加上各种测试之间的界限其实很模糊，这就造成了很多人理解上的混乱。

比如大家在工作讨论时，经常说做性能测试，但对于做什么样的“性能测试”，每个人有不同的看法，而且又经常表达不清。这就造成来交流不畅，甚至是误解，从而严重地影响了工

作的速度。我见过很多次因为对性能测试定义和交代不清，造成了老板和员工之间 / 员工和员工之间的理解误差。

性能测试的种类颇多，各自有不同的测试目的、测试环境、负载等等；这里面最重要的是测试目的和负载的大小变化。我们这一讲就一起来分一下类。

性能测试的分类方式

性能测试如何分类呢？我们需要从几个方面来看，包括测试目的、测试环境、负载流量、测试对象、负载数据、黑盒白盒等。

测试目的

测试目的是最重要的方面。大体上有几种目的：

1. 测量服务速度（Speed）：确定程序是否能够快速地响应用户的请求，这个服务速度一般包括延迟和吞吐率两个指标。速度通常是应用程序最重要的属性之一，因为运行缓慢的应用程序容易丢失用户。
2. 测量可扩展性（Scalability）：确定应用程序是否可以在用户负载和客户流量增大情况下还能正常地运行。
3. 测量稳定性（Stability）：确定在各种极端和恶劣环境下，应用程序是否能稳定运行。
4. 测量性能瓶颈（Performance Bottleneck）：性能瓶颈是应用程序和系统中的最影响整体性能的因素。瓶颈是指某个资源不足而导致某些负载下的性能降低。一些常见的性能瓶颈是 CPU、内存、网络、存储等。

测试环境

性能测试的环境也有几种，主要是开发环境还是生产环境。开发环境里面更多的是简单的测试来发现一些明显问题，而生产环境测试一般是开发环境测试通过后才进行的。

负载流量

根据测试的负载大小来分：是小流量，正常流量，还是超大流量。除了大小，负载变化的速度也需要考虑。

测试对象

测试的对象可以是只针对一个代码功能，或是整个代码模块，亦或是整个系统。

负载数据

测试的负载数据可以是真正的生产环境中的请求和数据，比如终端客户的网站请求和上传数据，也可以是人工模拟出来的请求及数据。

黑盒白盒

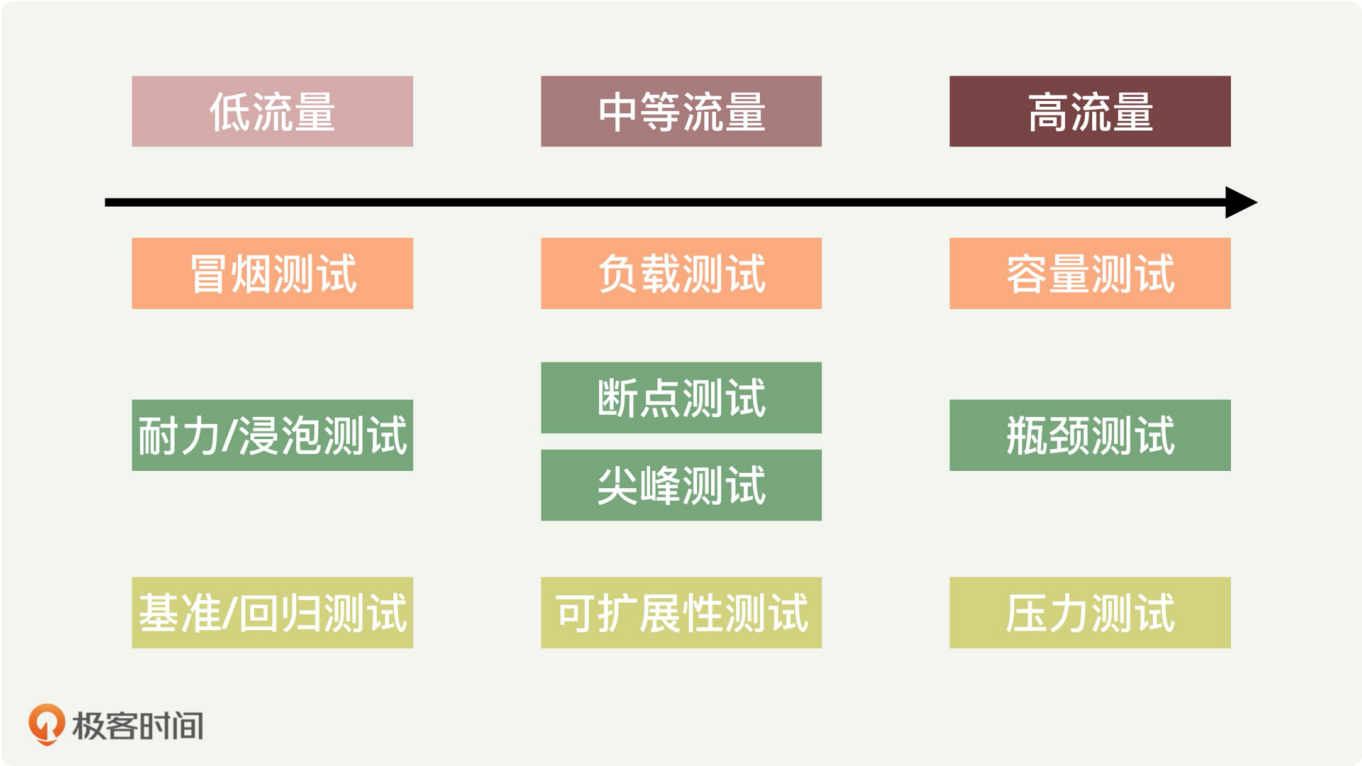
如果把被测试的对象当作一个整体，不关心它的内部工作机理，也就是把它当作一个黑盒子，那么这种测试就是黑盒测试。反之，如果你也关心它的内部构件和内部设计，就是把它当作白盒子来测试。

性能测试的种类

我们现在看看你可能经常会提到的各种测试，包括负载测试、容量测试、压力测试、断点测试、瓶颈测试、尖峰测试、耐力测试、基准测试、可扩展性测试和冒烟测试这 10 种。我分别说说它们是什么特点，尽量用刚刚讲过的分类方式归归类，并且适当举例说明。

需要说明的是，业界对于不同的测试类别其实也是各说各道，没有特别统一而严格的定义。我也是尽我所能，根据我的经验和实践来帮你理一理。

因为性能测试的种类多，我尽量把它们按照某种方式归类一下，帮助你理解和记忆。大体上可以按照负载流量的大小分成三类：低流量、中等流量和高流量。这里的流量高低是相对于生产环境中的流量而言的。当然，它们的实际界限其实很模糊。



冒烟测试（Smoke Testing）

冒烟测试是开发人员在开发环境里执行的简单测试，以确定新的程序代码不出故障。冒烟测试目的是确认系统和程序基本功能正常。冒烟测试的执行者往往就是开发人员，但有时也让运维人员参与。

耐力测试（Endurance Testing）和浸泡测试（Soak Testing）

耐力测试（或者耐久测试）有时也叫浸泡测试，是一种非功能性测试。耐力测试是长时间测试具有预期负载量的系统，以验证系统的行为是否正常。举一个例子，假设系统设计工作时间是 3 小时；我们可以对这一系统进行超过 3 小时的测试，比如持续 6 小时的耐力测试，以检查系统的耐久性。

执行耐力测试最常见的用例是暴露某些不易重现的问题，如内存问题、系统故障或其他随机问题。

这里的偏重点是**测试时间**，因为有些程序和系统的问题只有在长期运行后才暴露出来。一个最明显的例子就是内存泄漏。一个程序或许短时间内运行正常，但是如果有内存泄漏，只要运行时间足够，就一定会暴露出这个问题。

基准测试（Benchmark Testing）/ 性能回归测试（Performance Regression Testing）

基准测试或者性能回归测试是着重“前后”对比的测试。

这种测试往往是开发过程的一部分，一般不需要具体的性能要求。代码的演化过程中经常需要确保新的代码不会对整个模块或系统的性能产生任何不好的影响。最简单的方法是对代码修改前后进行基准测试，并比较前后的性能结果。执行基准测试的重点是**保证前后测试环境的一致**，比如负载流量的特征和大小。

负载测试（Load Testing）

负载测试用于验证被测试系统或者程序是否可以处理预期的负载流量，并验证正常和峰值负载条件下的系统和程序行为。这里的负载可以是真正的客户请求，也可以是仿真的人工产生的负载。

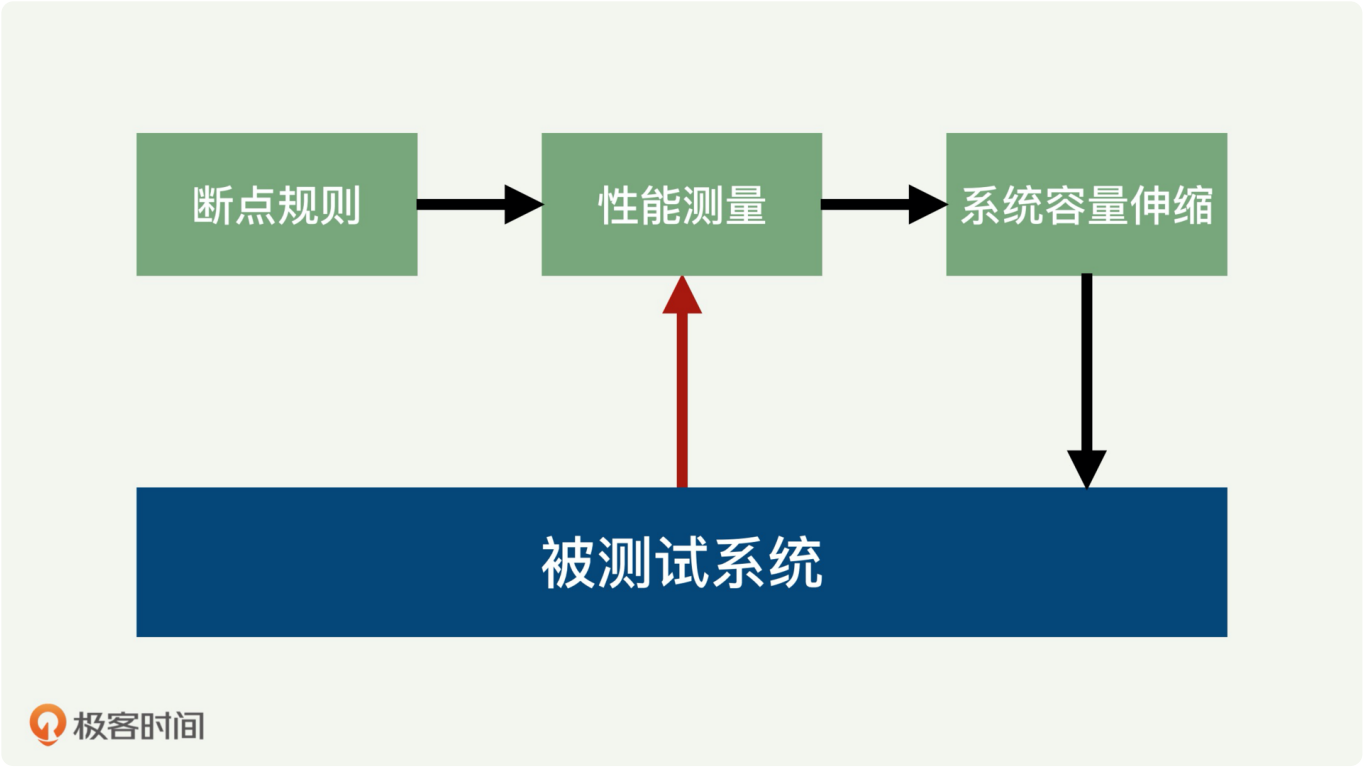
我认为负载测试的定义有时太广泛和模糊，很多其他测试都可以看作是负载测试的一种，比如马上就要讲到的容量测试，其实就是一种负载测试。

断点测试（Breakpoint Testing）

断点测试类似于压力测试或者容量测试。这种测试的过程是随着时间的推移而增大流量负载，同时监视系统的预定故障条件。

断点测试也可以用来确定系统将达到其所需规范或服务水平协议的最大容量，并且自动采取措施来纠正或者缓解。比如云计算环境中，我们可以设置某种性能断点，用它们来驱动某种扩展和伸缩策略。

比如一种性能断点可以是根据用户的访问延迟。如果延迟性能测量的结果是已经超过预定的阈值，就自动进行系统容量调整，比如增加云计算的服务器。反之，系统容量也可以根据断点的规则来减少，以节省成本。如下图所示。



尖峰测试（Spike Testing）

尖峰测试用于确定系统在负载（比如用户请求数）突然变化时的系统行为。这种测试是通过突然增加或减少由用户产生的负载来观察系统的行为。

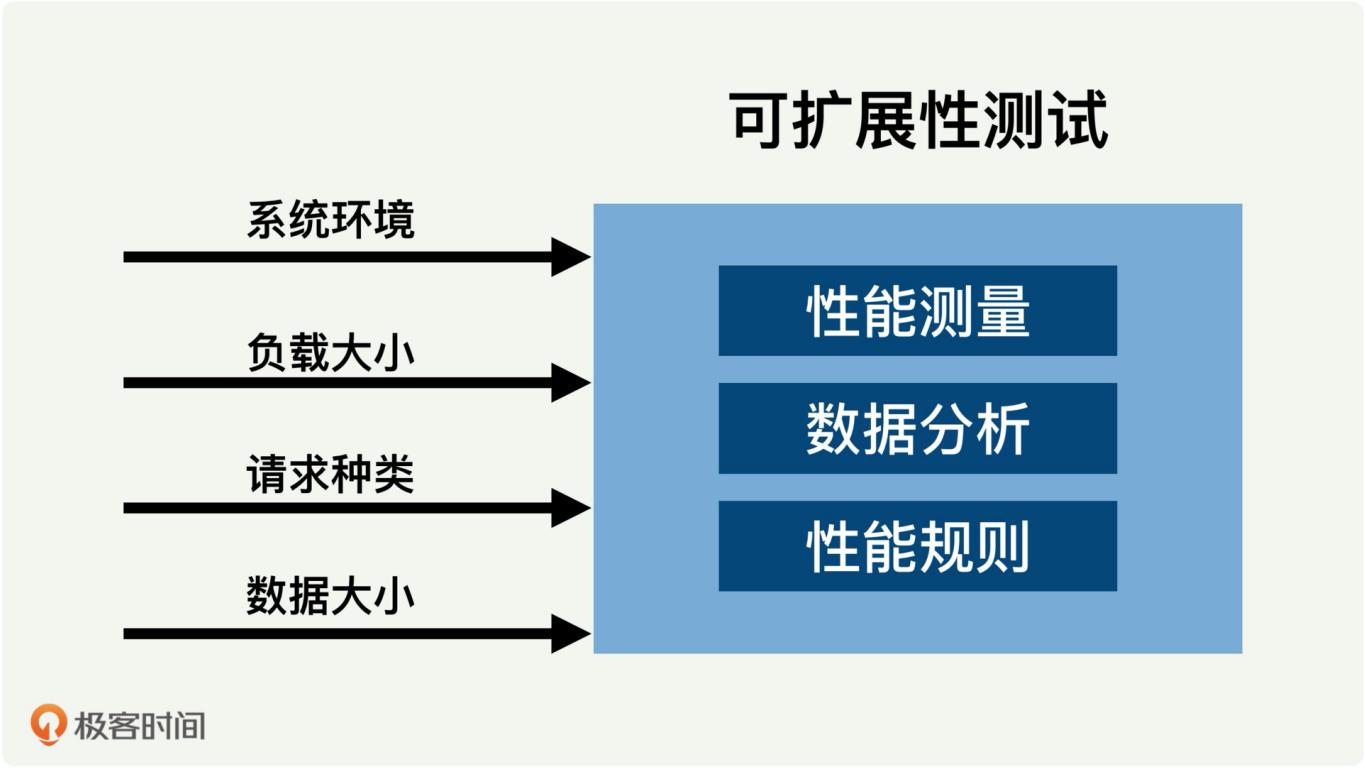
测试的目标是确定性能在这样的场景下是否会受损，系统是否会失败，或者是否能够处理负载的显著变化。尖峰测试的核心是**负载变化的突然性**，所以也算是一种压力测试。

可扩展性测试（Scalability Testing）

可扩展性（或者叫可伸缩性）测试用于确定一个程序和系统的非功能性特征能不能在变化的环境里合理扩展。这里的环境变化包括系统环境的变化、负载量的大小、请求的多样性、数据量的大小等。

在系统环境变化时，同步的测量和观察各种性能指标，并进行数据的分析，从而确定在各种环境下被测试系统的可扩展性。如下图所示。

这个测试的主要目的是了解系统在什么样的环境中，以及什么样的变化会导致系统不能扩展。发现这些环境后，可以进一步有针对性的分析和加强。



容量测试（Capacity Testing）

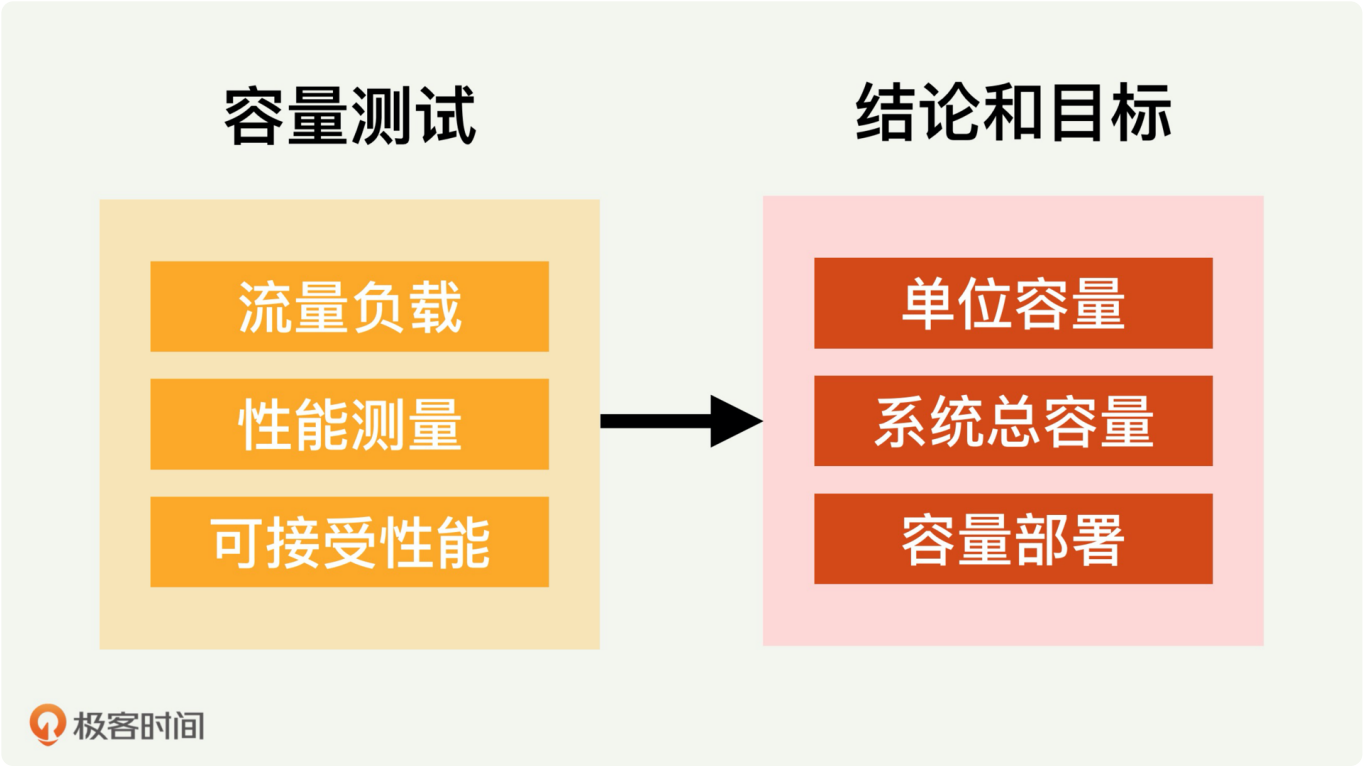
容量测试（或者叫体积测试，Volume Testing）是用于确定一个单位容量能够支持的最大负载。比如一个程序运行在某种服务器上，我们有时需要知道每台服务器能够支持的最大负载（例如客户数），从而决定需要部署多少台服务器才能满足预定的总负载要求。

容量测试一般是会不断增大负载，并且不断地测量各种性能指标。在性能目标变得不可接受之前，系统和程序可以成功处理的负载大小，就是单位容量可以承担的负载。为了尽量让得到的结果匹配实际生产环境，采用的负载流量最好是真正的生产环境的请求和数据。

正常生产环境中的流量和数据或许不够大到让一台服务器超载，因此我们需要解决这个问题。很多公司的解决方案是把其他服务器上的请求重定向到某一台被测试服务器，从而让这台服务器适度超载。这种机制我后面会用一讲专门讨论。

容量测试是确保系统稳定的重要一环。只有进行彻底的容量测试，并有相对应问题的解决方案，才可以使我们能够避免将来出现潜在的超载问题，例如增加的用户数或增加的数据量。

如下图所示，容量测试至少包含三个部分：可调节的流量负载、性能的测量、可以接受的性能指标。这三个部分一起就可以决定单位容量（比如一台服务器）的最大负载容量。这个数据可以帮助我们做各种决策，包括预估系统能负担的总负载；或者根据预期负载来决定部署多少台服务器。



瓶颈测试（Bottleneck Testing）

瓶颈测试其实可以看作一种特殊的压力测试。它的目的是找到被测试系统和程序的最制约的资源类型（比如 CPU 或者存储）。瓶颈测试并不局限于只找到最制约的一个瓶颈，它也可以同时找多个性能瓶颈。

找多个性能瓶颈的的意义主要有两点：

1. 如果最制约的瓶颈资源解决了，那么其他制约资源类型就自动会成为下一个瓶颈，所以需要未雨绸缪。
2. 系统设计时可以考虑在几个资源之间做些平衡，比如用内存空间来换取 CPU 资源的使用。

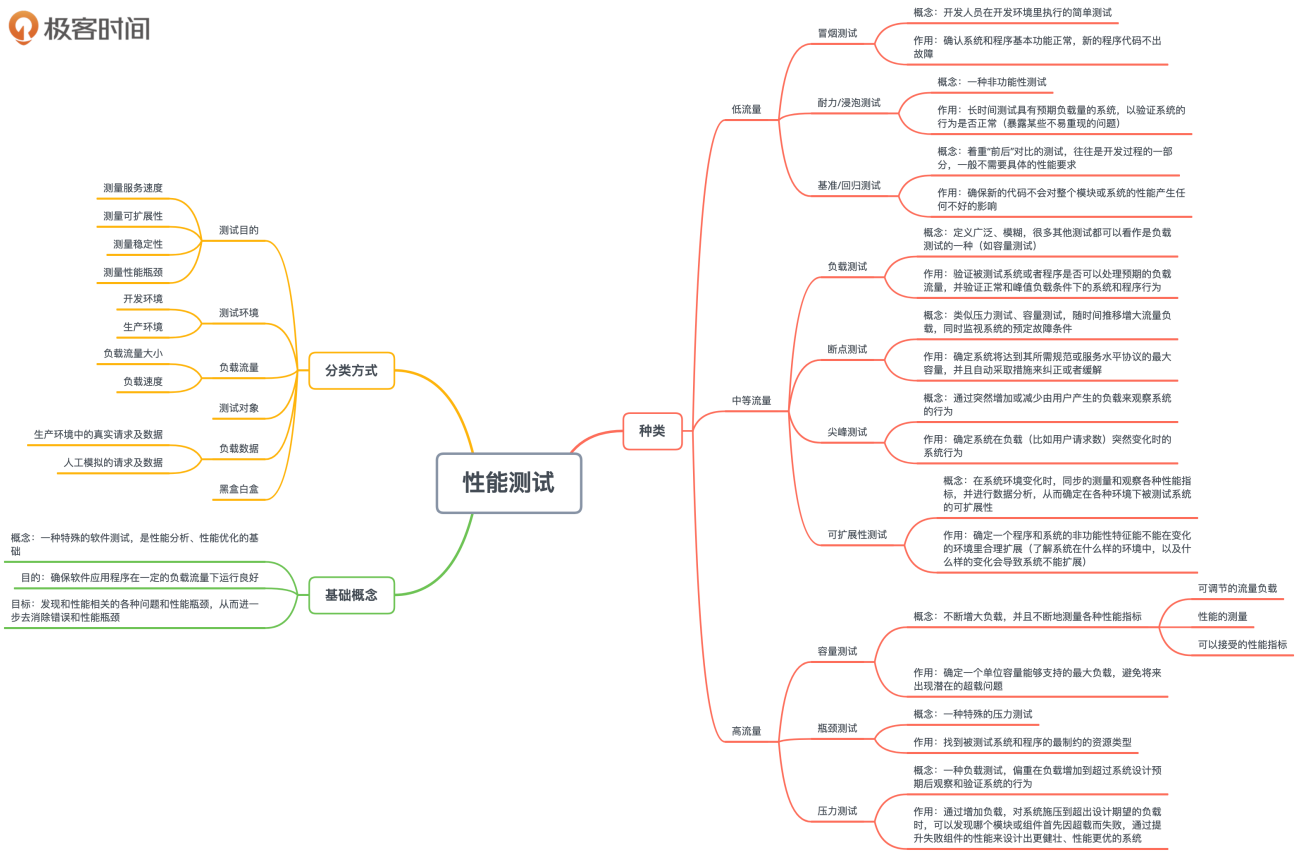
压力测试（Stress Testing）

压力测试也是一种负载测试，不过它偏重的是在负载增加到超过系统设计预期后观察和验证系统的行为。当我们通过增加负载，对系统施压到超出设计期望的负载时，就能发现哪个模块或组件首先因超载而失败。这样我们就可以通过提升失败组件的性能来设计出更健壮、性能更优的系统。

相对于容量测试，压力测试的目的是为了暴露系统的问题，因此采用的负载不一定是真正的生产数据和客户请求。

总结

这一讲重点讲了几种性能测试。



在工作中和别人交流时，你一定还会听到各种不同叫法的性能测试。我从业多年的总体感觉就是，性能测试的种类太多，甚至对某一种测试怎么进行也众口纷纭。给人的感觉，就像古诗里面所说的，“乱花渐欲迷人眼”。

虽然各种性能测试叫法不一，但万变不离其宗，你只要主要抓住几点就行，比如分类的方式，流量的大小和测试的目的。

希望通过本讲的讨论，你能对不同的性能测试之间的区别更清楚一些了，以后工作交流和阅读文献时能搞清楚它们是什么种类的测试，从而对症下药，做好测试规划和合理的分析。

思考题

今天讲了这么多种测试，你平时用过几种？

对于没有用过的测试种类，你清楚了解它们的使用场景吗？

如果有的测试种类看起来还挺有用但自己没有试过，尝试着设计一个，并在工作中表现一下，把这种测试和其他已经有的测试种类的区别好好分析一下，让同事们和领导知道你才是性能测试专家！

欢迎你在留言区分享自己的思考，与我和其他同学一起讨论，也欢迎你把文章分享给自己的朋友。

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇　08 | 经验总结：必须熟记的一组常用性能数字

下一篇　10 | 性能测试的规划和步骤：为什么性能测试不容易一蹴而就呢？

精选留言 (2)

写留言

张智凯
2019-12-18

没做过测试，在做运维，期待后续的讲解

展开

Q

2019-12-17

关于测试方法这块确实了解的比较少，工作中用到的最多就是负载测试，对机器做过载测试。